NOAA NESDIS
Central Satellite Data Processing Center

Concept of Operations for the Archive, Access and Distribution System As part of Comprehensive Large Array-data Steward-ship System (CLASS)

(Version 1.0)

October 25, 2002

Document Review History

Reviewer	Version Reviewed	Signature	Date
Constantino Cremidis/CSC	1.0		01-Oct-02
Alexander Kidd/OSDPD	1.0		01-Oct-02
Geof Goodrum/NCDC	1.0		01-Oct-02
Ted Habermann/NGDC	1.0		01-Oct-02
Eric Kihn/NGDC	1.0		01-Oct-02
Carlos Martinez/TMC	1.0		01-Oct-02
David Vercelli/NESDIS	1.0		01-Oct-02

Table of Contents

Table of Contents ..... **i**
Table of Figures ..... **ii**
Section 1 Introduction **1**
1.1 Scope 1
1.2 Applicable Documents 1
1.3 Definitions 2
Section 2 System Overview **4**
2.1 System Purpose and Scope 4
2.2 System Background and Context 4
Section 3 Management Policy **5**
3.1 Management 5
3.2 Governing Policies 7
Archiving 8
Preservation 8
Distribution 8
Operation 8
3.3 System Constraints 9
3.4 System Risks 9
Section 4 System Description **10**
4.1 Top-Level System View 10
4.2 Functional and Operational Capabilities 14
Operation 15
Archiving 15
Preservation 15
Distribution 16
External Interfaces 16
4.3 System Characteristics 16
4.4 Reference Architecture 18
4.5 System Interfaces 19
Section 5 System Environment **20**
5.1 Organizational Environment 20
5.2 Operational Environment 20
5.3 User Environment 20
5.4 Development Environment 21
Section 6 System Operations **21**
6.1 Operational Description 21
6.2 Significant Operational Requirements 21
6.3 Operational Interfaces 21
6.4 Operational Scenarios 22
6.5 Personnel Requirements 28
List of Acronyms **29**

Table of Figures

Figure 1 Management Organization 6
Figure 2 Process Overview of the Archive, Access and Distribution System 11
Figure 3 Top-level functional view 14
Figure 4 Reference Architecture: Current and Proposed 18
Figure 5 - Ingest - Scenario 1 22
Figure 6 - Order Placement - Scenario 2 24
Figure 7 - Delivery - Scenario 3 26
Figure 8 - Customer support - Scenario 4 27

Section 1 Introduction

This document covers the concept of operations for the Archive, Access and Distribution System, which is part of the National Oceanic and Atmospheric Administration (NOAA) Comprehensive Large Array-data Stewardship System (CLASS). CLASS, and consequently the Archive, Access and Distribution System are evolving, as new campaigns become operational. The Archive, Access and Distribution System needs to extend its resources and capabilities to satisfy the requirements for new data types.

The Archive, Access and Distribution System Concept of Operations covers the design, development, integration, operation, maintenance, and management policy for the lifecycle of the evolving project.

This document is based largely on experience gained through the development of the Satellite Active Archive (SAA), which is a successful implementation of the capabilities needed today; a user interface, automatic ingest, online search, automatic delivery, and special services such as subscription orders, off-line search and order, and push delivery.

1.1 Scope

In this document, the Archive, Access and Distribution component of CLASS is often referred to as simply "the system." This document describes the system from five points of view, as follows:

System Overview, Section 2, describes the scope of the intended use of the system and provides background information.

Management Policy, Section 3, describes the system goals, governing policies, constraints, and risks.

System Description, Section 4, describes the system capabilities and unique characteristics.

*System Environmen*t, Section 5, describes the operational, user, and development environments for the system.

System Operations, Section 6, describes the system operations through a set of scenarios. The scenarios identify the requirements that have a major impact on operations. The scenario descriptions provide a view of the anticipated system hardware operation, software, management operations, and the users.

1.2 Applicable Documents

This section presents the applicable documents.

Title	Reference Number
NOAA/NESDIS Enterprise Information Technology Architecture Plan	
NOAA Archive and Access Architecture	
CLASS System Architecture (July 2001)	
Satellite Active Archive system description	http://library.saa.noaa.gov
NESDIS List of IJPS Day One Products	
Polar-orbiting Operational Environmental Satellite Ground Segment Upgrade Description & Requirements For Initial Joint Polar Satellite System	NO-IJ/OSD-99-0005-ROUO
Polar-orbiting Operational Environmental Satellite System Requirements for Initial Joint Polar-orbiting Operational Environmental Satellite System (RDN4)	
Data Denial Implementation Plan (DDIP) for The Co-operation between NOAA and EUMETSAT on an Initial Joint Polar-orbiting Operational Environmental Satellite System (IJPS) Agreement	
NPP Mission System and Operations Concept (Volume 1)	GSFC 429_99_02_02
NPP Mission System and Operations Concept Data Dictionary (Volume 2)	GSFC 429_99_02_02
NPP Mission Requirements Specification, Volume 1, (Level 2)	GSFC 429_99_02_03
NPP Mission Requirements Specification, Volume 2, (Level 2)	GSFC 429_99_02_03

1.3 Definitions

Campaign
Scientific endeavor encompassing data from a number of single data sources.

Data set

A collection of related data that is ingested as a single unit. A data set may consist of one or more files that contain data and metadata or one or more related database tables.

Ingest
To take in for storage, create an inventory entry and possibly, depending on the data type, extract and store data from which visualization products can be generated. Some visualization products are used as an aid to data selection, others are final products that can be used for data analysis. Also, performs basic quality controls and integrity of data from as defined by agreement between NESDIS and the data provider

Inventory Catalog
A catalog that contains the unique description of all ingested data-sets.

Large-array
In the context of the Archive, Access and Distribution System, large-array data refers to remotely-sensed enviornmental data from any of the campaigns that are included in this system. Characteristics of the sensor include continous measurement that is recorded and subdivided into measureable blocks covering finite time intervals.

High Availability
For Archive, Access and Distribution System, high availability refers to duplicate capabilities on equivalent hardware and software systems, at two separate sites with dedicated high bandwidth connections. Between the two systems 24x7 operation and 99.9% availablity.

Off-line Search and order
An order submitted by an operator in accordance with user-defined search criteria. All data sets that match the search criteria are automatically submitted as part of such order. Off-line orders, also known as bulk orders, are processed at low priority and are for data that spans a large period of time.

Subscription order
An order placed automatically for a newly ingested data set that meets user-defined criteria. These criteria, along with user identification and delivery intructions, constitue a subscription.

Visualization Products
Views generated under the control of the Archive, Access and Distribution System that show the content of data sets. The products may provide views of each data set for a data type or samples of products.

Visualization tools
Tools that generate visualiztion products.

Section 2 System Overview

The mission of the Archive, Access and Distribution System is to manage the archive, access and distribution of data for CLASS. The blend of organizational policies and technology must adapt to the campaigns, the users, and the commitment of the funding agency. This document describes the scope of the processes and capabilities that are needed in terms of four main categories: operation, preservation, archiving, and distribution.

2.1 System Purpose and Scope

The Archive, Access and Distribution System is a component of NOAA CLASS that provides the capabilities to archive and distribute large array-data for the following seven campaigns and including new *in-situ* and remotely sensed data:

1) POES, including DMSP (Polar-orbiting Environmental Satellites for NOAA and DOD, Defense Meteorological Satellite Program)
2) GOES (Geostationary Operational Environmental Satellites)
3) NEXRAD (NEXt generation weather RADAR)
4) EOS/MODIS (Earth Observing System/Modern Operations Distributed Infrared Spectrometer)
5) NPP (NPOESS Preparatory Project)
6) METOP (European Meteorological Operational Satellite)
7) NPOESS (National Polar-Orbiting Operational Environmental Satellite System)

The Archive, Access and Distribution System development and deployment relies on the lessons learned from the operation of the successful SAA. The SAA is the baseline system. The SAA requires upgrades to manage the expected growth in these campaigns. For the most part, upgrades will extend existing capabilities. Additional development will implement requirements for supporting distributed operations, handling new types of data and providing high availability.

The SAA is an operational component of the Information Processing Division (IPD) of the NOAA Office of Satellite Data Processing and Distribution (OSDPD). The SAA distributes data only in electronic form. Offline or cartridge delivery is provided through arrangement with NOAA's National Climatic Data Center (NCDC). OSDPD and NCDC are line organizations of the National Environmental Satellite, Data, and Information Service (NESDIS). CLASS is administered by NESDIS.

2.2 System Background and Context

CLASS is being developed by the NESDIS, the NOAA line office that operates the United States' Geo-stationary Operational Environmental Satellites (GOES) and the Polar-orbiting Operational Environmental Satellites (POES). The NESDIS mission is to ensure timely

access to global environmental data and to provide information services, including Earth system monitoring and assessments of the environment, in order to promote, protect, and enhance the Nation's economy, security, environment, and quality of life. NESDIS operates the National Data Centers (NNDC) and the SAA as part of this mission. Together, these groups manage one of the largest collections of atmospheric, geophysical, and oceanographic data in the world.

Currently, satellite and in-situ sensor data sets are archived, managed, and distributed by different centers in NESDIS. The approach has met previous needs, however the Centers will soon be overwhelmed by the increasing volume and complexity of data that they are responsible for. By the year 2015, current plans for the campaigns of NEXRAD, GOES, POES, METOP, NPP/NPOESS, and EOS/MODIS, will increase the volume of data from less than 1 to more than 14 petabytes. CLASS is being carried out to ensure that NESDIS can continue to provide services in the face of that increase. This document describes the concept of operations to fulfill the data archive, access and distribution aspect of the CLASS mission.

Section 3 Management Policy

3.1 Management

Multiple departments of NOAA contribute to the CLASS design process. The technical management team includes members from the National Climatic Data Center (NCDC), the National Geophysical Data Center (NGDC), the Office of Satellite Data Processing and Distribution (OSDPD), and NESDIS headquarters. .

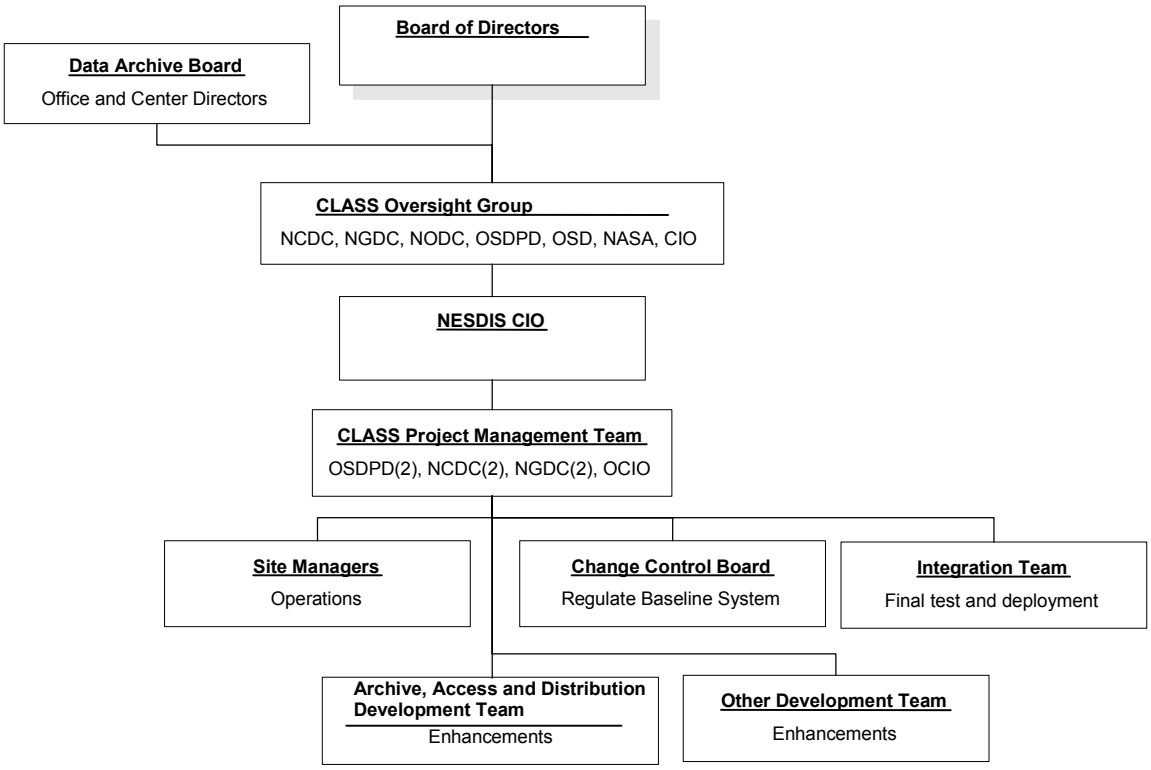

Figure 1 Management Organization

Board of Directors
Provides high-level review and sets priorities for the development of the CLASS project. Sets data set and product priorities and resolves issues brought to them by the CLASS Oversight Group.

Data Archive Board
Determines the data types that will be included in the CLASS project.

CLASS Oversight Group (COG)
Provides management review and direction to the CLASS Project Management Team (CPMT). Ensures management direction is being followed. Resolves or pushes up any policy issues or tasking issues brought to them by the CPMT. CPMT reports to the CIO and this group.

NESDIS CIO
Provides strategic oversight and direction to CPMT. Resolves technical issues and ensures enterprise focus and meets with CPMT as needed.

CLASS Project Management Team (CPMT)
Plans for enhancements, schedules activities, recommends polices to the COG and implements approved policies for operations, development, implementation, and deployment.

CPMT assesses the success of policies that affect operations and, as necessary, determines improvement.

Site Managers
Carry out policies set by the CPMT and coordinates operator activities.

Change Control Board (CCB)
Reviews change requests for new development, determines priority and readiness of the modifications to be made to the system.

Integration Team
Performs integration and deployment activities.

Archive, Access and Distribution Development Team
Performs development and unit testing for the Archive, Access and Distribution System.

Other Development Teams
Performs development and unit testing for other sub-systems of CLASS.
System Engineering Team (SET) ensures the technical integrity of the system.
Software Engineering Process Group (SEPG) promotes continuous process improvement.

3.2 Governing Policies

The detailed policies for lifecycle management must be defined by the CPMT. The general organizational objectives are as follows:

- Development of software is geographically distributed, but controlled by a centralized Configuration Control Board.
- Installation of the system will be at two sites, leveraging existing facilities.
- Operations for archive and distribution of data sets is fully automatic and depends upon electronic transfer for source data.
- Operators are able to conduct supporting activities remotely, using a secure interface. Examples of supporting activities are customer help, reconfiguration of hardware for load balancing, and data migration. Operators coordinate with the housing facility, but system administration of hardware is the responsibility of the housing facility site management.
- Unless a disaster contingency plan is in progress, each housing facility must have one operational system.

The policies cover the following four categories of capabilities and responsibilities:

1) Archiving - take in data, catalog, move to archive
2) Preservation - backup, duplication, and migration
3) Distribution - Internet-interface, user access, order fulfillment
4) Operation - people managing the system, and policies.

Archiving

Primary data sources are limited to NOAA, NASA, and DoD institutions that require long-term preservation and distribution of non-classified data.

The Archive, Access and Distribution System flexibility must be sufficiently robust to cover all the needs of the seven sets of large-array data. Other sources of data (e.g., *in-situ*) may be added as well to the system, by application to the Data Archive Board and the Board's acceptance.

Any data set must conform to certain requirements on content, format, and handling as defined by agreement between NESDIS and the data provider to be acceptable to the system, but some changes in the system may be required to accommodate each new data set. Based upon direction from the Data Archive Board and/or CLASS Oversight Group, the CPMT initiates analyses and conducts reviews to assess the compatibility of each new data set with the Archive, Access and Distribution System, and determines the system changes required to enable the acceptance of this data set.

Preservation

The System provides the ability to preserve data from loss. Minimally, preservation requires that two copies of a data be kept at all times. A minimum of 50 miles must separate the storage locations of those copies.

Distribution

Automatic distribution of electronically transferable data sets is the preferred method of delivery; the system will be able to provide offline delivery of data. Pricing is not the responsibility of the Archive, Access and Distribution System. The system supports an interface to an Order Management System (OMS) and provides necessary and sufficient information to the OMS so that the OMS may calculate pricing and carry out charging activities.

Operation

The management team publishes and enforces policies for the operations of the resources in normal and contingency modes of operation. The team provides plans to accommodate the continuous growth of data sets and types of data sets that the system must support.

3.3 System Constraints

System constraints are system requirements that are imposed on the Archive, Access and Distribution System by NESDIS and the CLASS System Architecture.

1. All financial operations, i.e. charging for data, must be conducted using an interface with the Order Management System.

2. Section 508 of the rehabilitation act requires that Federal agencies electronic and information technology is accessible to people with disabilities.

3. NESDIS data archive policies must be followed.

4. The Archive, Access and Distribution System shall operate at two sites at least 50 miles apart, one site managed by OSDPD, the other by NCDC.

5. The external network bandwidth capacity at each site needs to be 150% of the expected ingest required by the combined sites.

6. Security shall be controlled at the data level.

3.4 System Risks

The overall CLASS system architecture is being defined. Therefore, there is a possibility that some requirements allocated to the data archive, access and distribution component may be inconsistent with the concept of operations described here.

Section 4 System Description

4.1 Top-Level System View

The system provides life cycle capabilities of archiving, distribution, preservation, and operation such that all approved campaign array-data may be preserved as defined by existing archive policy and distributed as requested to customers and for disaster recovery. The scope of these capabilities includes the ability to scale system functionality to continuous growth in campaigns and the preservation needs of the data. The capabilities are defined below:

Archiving - Archiving encompasses all activities to obtain the data, verify the data sets, inventory the data-sets, and place the data into a near-line archive system. The descriptive data stored in the inventory supports user-initiated searches. Automatic background processes synchronize the archive content at both CLASS Sites, and annotate the inventory catalog that the data set is stored in both archives.

Distribution: Distribution encompasses all activities that provide for Internet enabled user access, data visualization, customer ordering, customer authorization, retrieval from near-line storage, and product generation. Customer authorization provides for both restricted data access, and charges for delivery of an order.

Subscription services provide specific customers with the ability to place standing orders. Standing orders provide the system with the information to automatically fulfill that order as a matching data set is ingested. The delivery of the order to the customer's site may be automatic as well, or the customer may be given a specific time period in which to obtain the order. In a similar manner, only specific customers are allowed to place orders for existing data by using an off-line search capability. An off-line search places an immediate order for any data that match the search criteria without further customer interaction.

Preservation: Preservation encompasses all operations that ensure the safety and availability of the data. The data must be backed up locally and remotely for disaster recovery. As the storage media degrades, data migration to new media ensures the integrity of the data.

Operation: Operation encompasses all system support activities. The CPMT is the head of all operations. The CPMT needs to plan enhancements, migration schedules, and resource usage, and define policy. Site managers need to provide for immediate maintenance, resource commitment, and adherence to policies within the site organization. Operational decisions require the detailed collection of performance and use statistics from the archiving, distribution, and preservation activities. The statistics are needed to generate clear concise reports to NESDIS management and the CPMT, site managers, system operators, and

customers to enable each group to understand the system operations from its point of view. System operators provide customer support through the electronic help desk, and carry out policy. The customers need to place orders easily, check the status of orders, and obtain help.

The Archive, Access and Distribution System overview in Figure 2 depicts the process flow that encompasses the vision of both the business and the technology components of the system. The roles played by the elements in this diagram are described below.

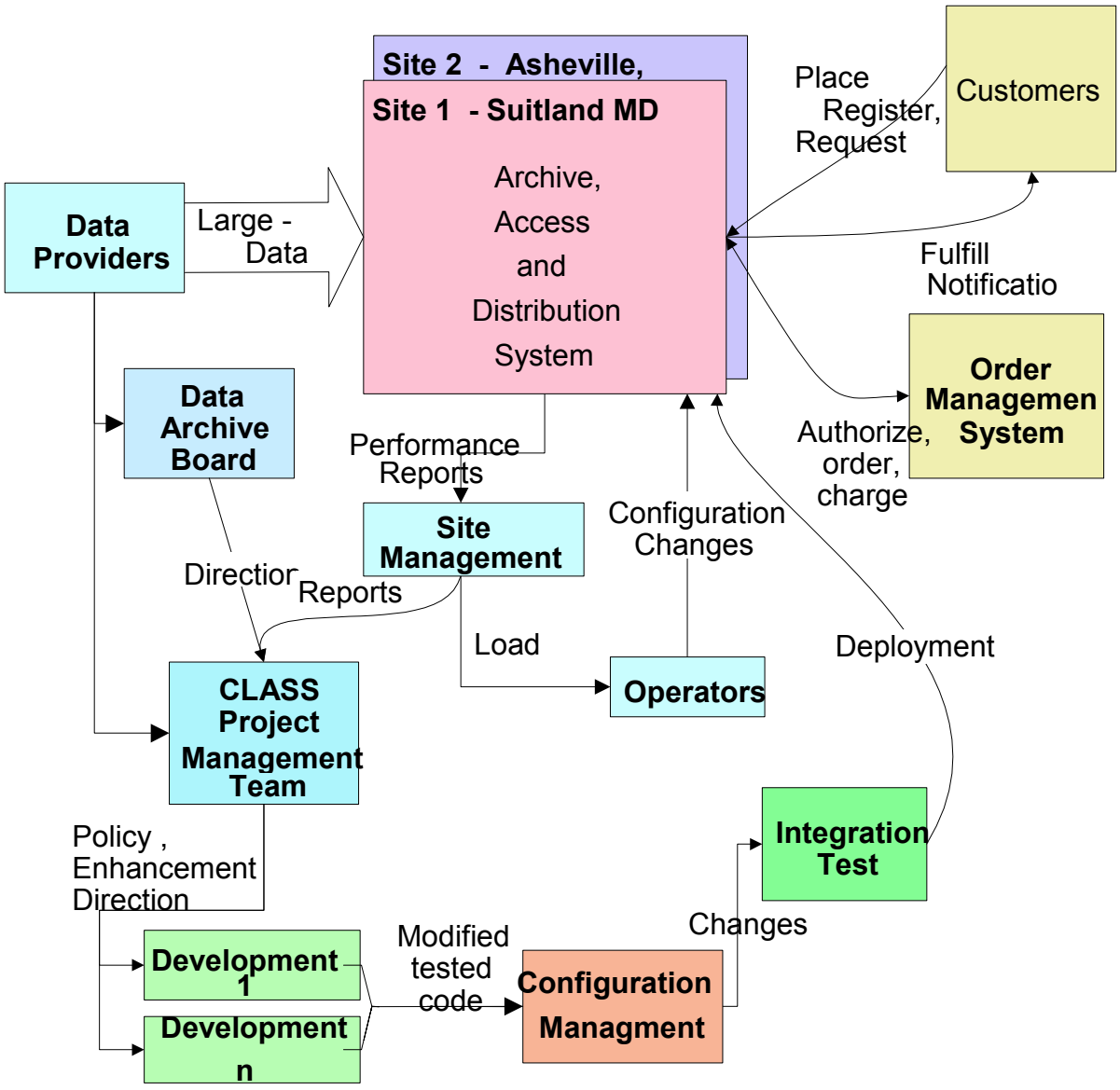

Figure 2 Process Overview of the Archive, Access and Distribution System

Sites
The system that performs the archiving, distribution, and preservation functions is deployed at two operational sites: Suitland MD, and Asheville NC.

Site Management
Site managers review the performance of the system. They receive performance reports that provide information on the utilization and performance of the system components under the processing load of ingest and distribution. Site managers assess changes in the system load and require operators to modify the system allocation of resources to fit fluctuations in demand.

Following the CPMT policies, the site management approves or denies customers for priority subscription services and automatic order delivery services. Because of the demand that can be placed on system resources, the site management decides which customers may receive priority privileges.

Operators
Operators carry out the day-to-day operations of the system; adjust configuration to support customer demand and provide electronic help desk support.

Data Archive Board
The Data Archive Board provides overall coordination and direction for inclusion of additional data types.

CLASS Project Management Team
The CPMT implements NESDIS and NOAA policies and prioritizes current projects. The CPMT defines the policies governing the privileges, and the site managers carry them out. The policies provide guidance to allow site managers to exercise complete control over site resource management.

The CPMT provides the design goals and the approval for development. Multiple development sites employ a unified configuration managed baseline. One site executes integration testing. On completion of a new baseline, the team deploys the new baseline to the different CLASS operational sites.

Development
Multiple development teams may contribute to a baseline.

Configuration Management
Configuration Management tracks change requests and verifies module level tests. Additional work may be required of the development team to correct problems that occur during system integration and test. A new baseline will be determined and frozen for integration test.

Integration Test

An integration team builds a new baseline and tests the new baseline. When successful, the new baseline is deployed.

Order Management System

The Archive, Access and Distribution System interacts with customers, and interfaces with an OMS. The OMS provides the billing and accounting service for customer transactions.

Data Providers

New products and new data sources will have to be approved by the Data Archive Board. The data providers make the data available to the Archive, Access and Distribution System by placing the data in a network location that can be accessed by a protocol that can execute automatically. Details of data provision processes must be designed between the system operational management, CPMT, and the data providers. Details of data provision include the acknowledgement and retransmission in case of a failure. New products and mew data sources will be approved by the Data Archive Board and the interface details will be defined by agreement between NESDIS and the data provider.

Customers

The Archive, Access and Distribution System provide at a minimum an Internet enabled interface to customers. Customers may search for data sets, browse sample images, and place orders for data. The system shall provide for electronic delivery of orders as well as physical media delivery. The interface to the OMS provides the avenue for approval of any orders that require customer funds for completion. Distribution of certain types of data can be restricted to certain users. Site management provides different priority privileges to customers on a limited basis. Customer privileges include subscription services, with either electronic transfer push or pull delivery, and bulk orders. Electronic pull delivery is the default for all delivery services

4.2 Functional and Operational Capabilities

The following diagram, Figure 3, shows the external interfaces of the Archive, Access and Distribution System to the data providers, customers, and OMS. It also shows the internal functions of the system, grouped into the four categories; Operation, Archiving, Preservation and Distribution.

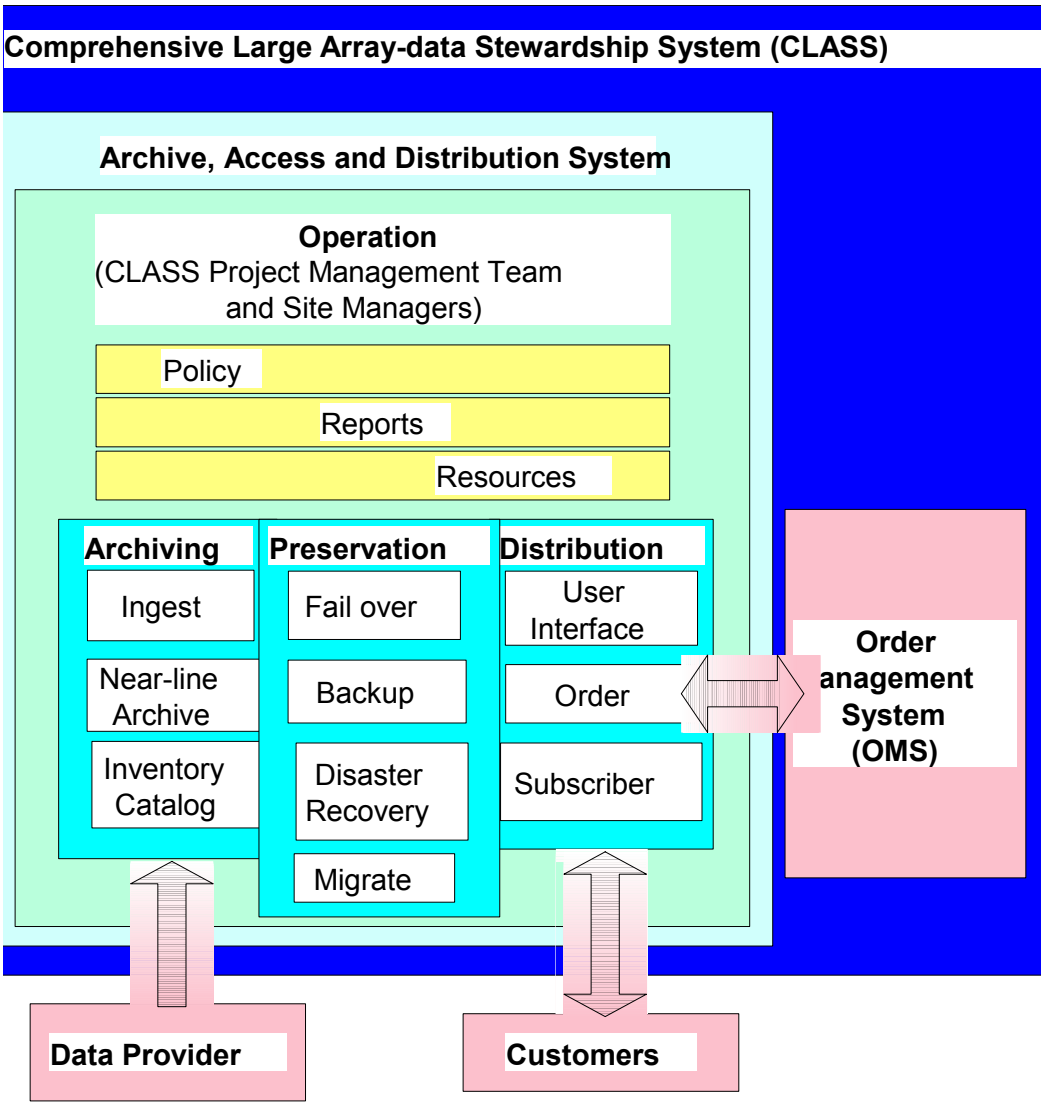

Figure 3 Top-level functional view

Operation

CPMT recommends and implements approved policies and procedures, and acts on analysis of system performance and resource allocation provided by system operators.

1) Policy – Policies and procedures must be set to define how system operations are conducted across all sites. The policies and procedures provide detailed guides and operational rules that must be followed by site managers and operators.

2) Reports – Effective, clear reports are essential for resource management, policy revisions, and for the extension of services to additional campaigns.

3) Resources – Resources are managed across the different operational sites, based on defined policies and procedures and system reports.

Archiving

Archiving provides all aspects of receiving the data from data providers and making data available to the rest of system.

1) Ingest - Ingest takes in the data from the data providers, develops or extracts descriptive data about the data set and in some cases, extracts and stores on-line data that can be used for visualization product generation. Ingest stores the descriptive data in the inventory catalog.

2) Near-line archive - The archived data set is stored in a robotic archive that allows for electronic retrieval.

3) Inventory Catalog - The inventory catalog contains data set entries that describe data sets archived in the near-line storage system. The inventory catalog is used to provide on-line real-time search services to a customer.

Preservation

Provides the abilities to keep the data over years and to maintain the quality of the data.

1) Fail-over – Fail-over provides the ability to switch between operational sites in the event that one site stops operation.

2) Backup - Local Backup provides duplication of the files in case one file is corrupted. Duplication of the archive provides a geographically distributed backup capability.

3) Disaster Recovery - In the case of catastrophic failures, the geographically distant files are used to bring the system back up to the current operational state.

4) Migrate - Migration of data is necessary as technology changes and as the media ages. Migration operations provide long-term data consistency.

Distribution

Distribution encompasses all necessary components to make data sets available to the user community. The user community may include other systems that provide other services, which employ the data.

1) User Interface - The user interface provides Internet access to users for the active archive services. The services provide for the ability to find the data, metadata, display data sets that match search criteria, view visualization products and order the data.

2) Order – Data sets and derivative products are received and delivered. Data products consist of any additional manipulation to a source data set. The data provider generates the product. Generally, this function provides the assembly of orders and the generation and delivery of requested products.

3) Subscriber - Special services are available to some customers. These customers have the ability to have orders executed as the data enters the system, or the ability to place off-line searches. An off-line search automatically creates an order for any data sets that match the criteria.

External Interfaces

There are six major external interfaces for the system.

1) Remote electronic data provider provides data files for ingest processing.

2) Customer accesses through a standard Internet interface.

3) Customer accesses help desk through email.

4) Customer receives delivery by electronic transfers.

5) The Archive, Access and Distribution system interfaces with the Order Management System for order authorization.

6) The Archive, Access and Distribution system interfaces with the Order Management System for media delivery.

4.3 System Characteristics

The Archive, Access and Distribution System, is be operational simultaneously at two sites, Suitland, MD and Asheville, NC. This configuration defines the operational characteristics of the system:

- Each site is a backup of the other, and as such, each has the processing, network and staffing to support full systems operations.

- Each site has a complete copy of the database.

- In normal operations mode, the two sites ingest different data flows. Data ingested at one site is replicated to the other site within four hours of ingest.

- Each data type has a preferred distribution site. The site that ingests the data is the preferred distribution site.

- When there is a large backlog of orders at one site, the distribution sub-system balances the order queue sending some orders to be processed at the other site, even if this is not the preferred distribution site.

- Visualization products created at ingest time are duplicated to the other site immediately after ingest. Visualization products are used by the web site to provide visual aid for data selection.

- Deploy-ability - The system must be easy to deploy.

In addition to these characteristics derived from the requirement for two operational sites, the system has the following operational characteristics:

- Reliability - The system must provide 99% dependable services for users.

- Extensibility - The system must support adaptation to changes in specification.

- Interoperability - The system must support the ability of exchanging information with data repositories outside CLASS.

- Scalability - The system must support growth to provide acceptable levels of 99% performance for new data flows.

- Recoverability - The system must provide error detection and recovery functions for all of its services.

- Efficiency - The system must perform its designated functions with minimum consumption of resources.

- Integrity - The system must provide protection against unauthorized access and modification.

- Security – The system must provide protection to customer’s privacy.

4.4 Reference Architecture

The reference architecture displayed in Figure 4 presents the current operational system and the enhancements that are required to accomplish the first full deployment of the Archive, Access and Distribution System. A brief description follows the diagram.

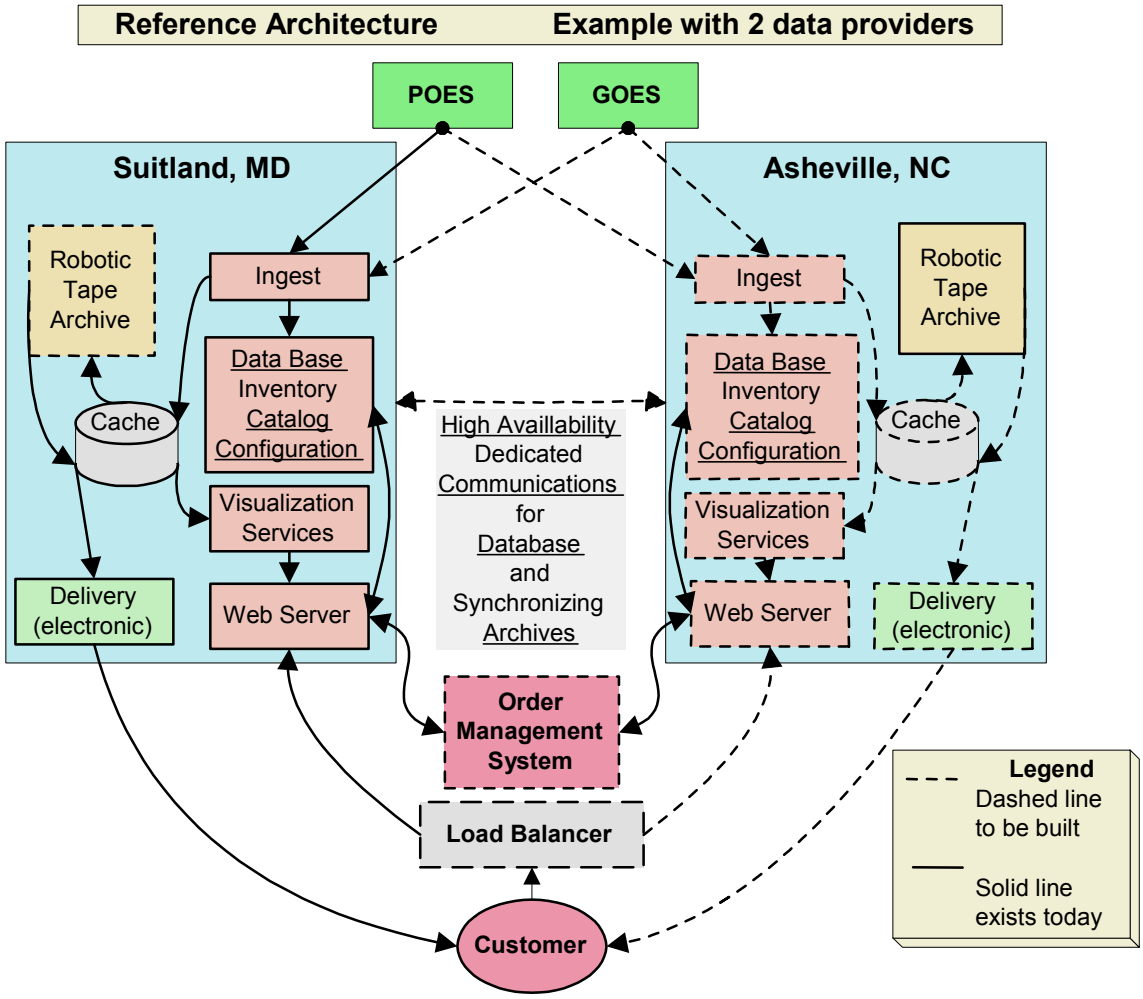

Figure 4 Reference Architecture: Current and Proposed

- The Suitland, MD facility needs additional archive capability and additional capability for disaster recovery. Although the components of SAA have not yet been enhanced for, the characteristics as described in Section 4.3, the Suitland capabilities are shown as operational.

- The Asheville, NC site has an archive but needs the installation of the baseline system and equipment. The Suitland facility needs a robotic tape archive.

- The software components ingest, delivery, visualization services and Internet services, are hosted on IBM equipment. The equipment needs to be upgraded to support the large array data sets based on the new campaigns.

- The SAA currently handles only POES data.

- When both sites are deployed, a mechanism that provides load balancing will be needed. In addition, the OMS is still in the process of being defined. The interface to the OMS will have to be developed.

4.5 System Interfaces

The system has six external interfaces. These are depicted in the scenario figures in Section 6 as indicated:

1) Remote electronic data provider provides data files for ingest processing.
   (Figure 5 - Ingest - Scenario 1)
2) Customer accesses through a standard Internet interface.
   (Figure 6 - Order Placement - Scenario 2)
3) Customer accesses help desk through email.
   (Figure 8 - Customer support - Scenario 4)
4) Customer receives delivery by electronic transfers.
   (Figure 7 - Delivery - Scenario 3)
5) The Archive, Access and Distribution system interfaces with the Order Management System for order authorization.
   (Figure 6 - Order Placement - Scenario 2)
6) The Archive, Access and Distribution system interfaces with the Order Management System for media delivery.
   (Figure 7 - Delivery - Scenario 3)

The system has three internal interfaces, as shown in Figure 4, Reference Architecture.

Dedicated communications between sites allow,

1) The robotic archives to duplicate data ingested at one site.
2) Synchronization of data and database content.
3) On-line access to selected cached data products.

Section 5 System Environment

5.1 Organizational Environment

The organizational environment spans several NESDIS data centers and offices. Each organization includes subcontractors needed to meet staffing requirements. The management policy (Section 3.2) in this document depicts the organizational chart and names the organizations involved in the Archive, Access and Distribution System and CLASS management.

5.2 Operational Environment

The Archive, Access and Distribution System use existing facilities as shown in Figure 4. The facilities are located in Asheville, NC and Suitland, MD. The facility administration supports the physical needs of the system operation. The system operators provide configuration definitions to the system to support current resource allocation needs. Operators have the capability to configure and monitor the operation of the system.

Each facility environment needs to have sufficient bandwidth to ingest all data if needed. Limits may be placed on distribution only if necessary due to the volume of user orders. Maintenance is conducted in the operational environment. Hardware components must be hot swappable to isolate maintenance and upgrades that can be performed by the system administration staff. Upgrades and maintenance to the system capabilities follow the policies that are defined by CPMT.

The processes within the system must be fully automatic for normal operations and for retries in the case of failed transmissions. Only when electronic communications are determined to be off-line or fail consistently should operators at CLASS sites communicate with other CLASS sites and data providers to determine a course of action to resolve the failure.

5.3 User Environment

Two user groups exist for the system. One group is the source data providers. The source data providers communicate electronically with the system. The Archive, Access and Distribution System communicate electronically with the source data provider for all normal operations. Interface rules must describe the processes to verify the correct ingest of data as defined by agreement between NESDIS and the data provider.

The second group of users order data. These customers employ current browser technology to place orders, or review status of an order. Email is the only direct means to communicate with the system operators for a typical customer. Special service users are supported and special services are granted by the site managers on a case-by-case basis according to the policies set by the CPMT. Users may have priority delivery methods such as push or pull subscriptions by electronic delivery.

5.4 Development Environment

Details for the development environment are described in the CLASS Master Project Plan.

Section 6 System Operations

6.1 Operational Description

The Archive, Access and Distribution System have two modes of operation: normal, and contingency. Normal operation includes three primary functions: archiving, distribution, and preservation. These functions are normally performed concurrently and automatically. Contingency modes with degraded phases functionality will be defined by the CPMT to address needs of disaster recovery, or maintenance.

The normal mode supports automatic, continuous ingest of source data into an archive, duplication to the alternate site, update to the inventory catalog, receipt of orders, fulfilling orders, distribution to customers, and customer profile management.

Data will be automatically copied from network attached or network accessible (on-line) storage media to the other CLASS site for disaster backup/recovery protection. As an intermediate step, the remote site may cache data to on-line digital media before processing to long-term storage. Process management information is exchanged automatically between the two sites.

The contingency mode addresses all other situations to support the best quality of service available. The contingency plans must provide the guidance for data migration, fail-over, disaster recovery, maintenance, and resource management. The resource management provides a necessary or negotiated quality of service that priority customers require.

Operators are only involved in implementing system configuration changes to support the contingency situations as required by the CPMT. Operators may also respond to customer email inquiries, and enter process improvement requests.

6.2 Significant Operational Requirements

Significant operational requirements are:

1) Two sites sharing archive ingest duties;
2) high availability (99.9%) of two sites;
3) full automation of archive and distribution;
4) full duplication of data at both sites.

6.3 Operational Interfaces

The system needs at least two types of operator roles: resource manager and customer liaison. As resource manager, the operator configures the system as needed to provide the quality of service that has been negotiated for a customer at a specific level of priority.

As customer liaison, the operator responds to customers helpdesk email requests. This operator receives problem reports, and can change priority of orders. Some status changes require approval of the site managers.

Only changes in customer status, or problem assessment, need to involve an operator. All normal functions of order placement are automated.

6.4 Operational Scenarios

1. Ingest

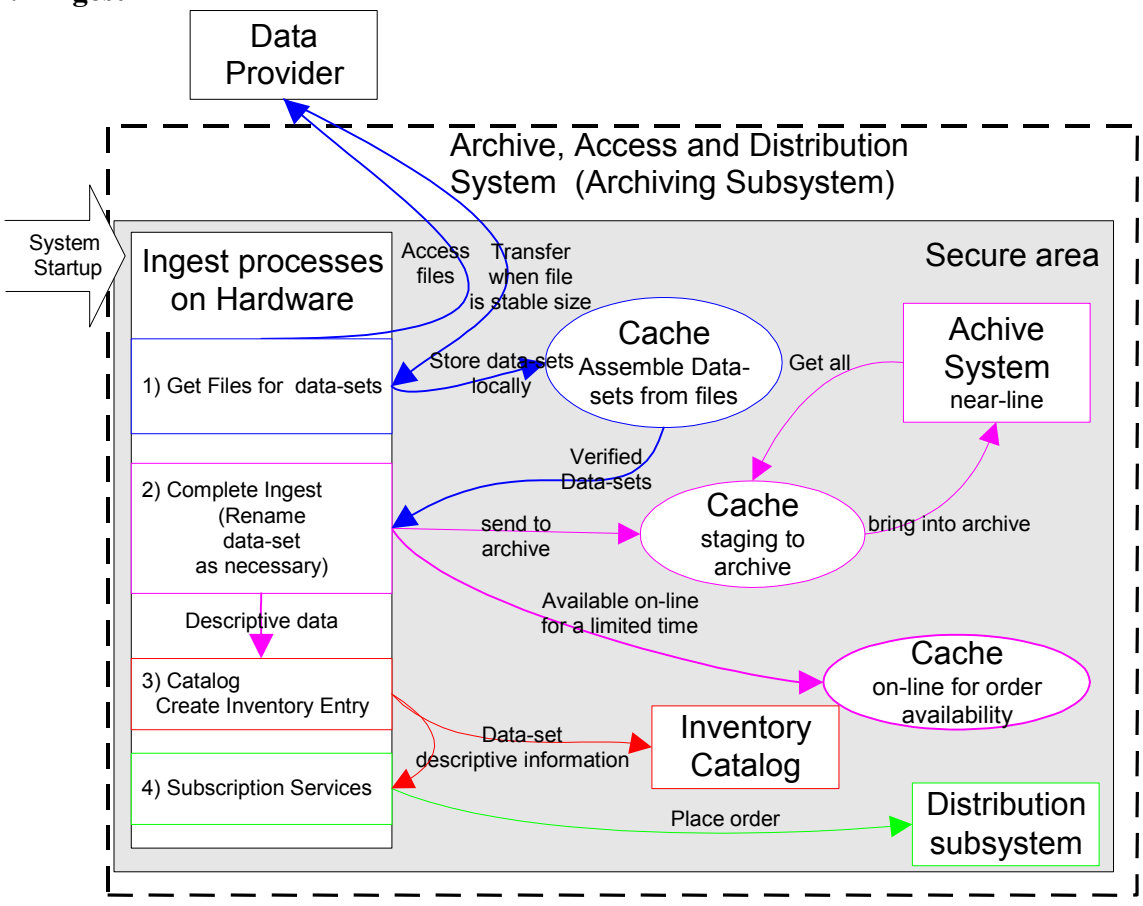

Figure 5 - Ingest - Scenario 1

The system configuration identifies the allocation of hardware to ingest processes and the network locations that provide cache reservoirs for source data and intermediate stages.

1. Archive, Access and Distribution ingest function locates data files from the data provider. Files are identified by both the location and the pattern of the file name. Before ingest transfers a source file into the system's secure area, it determines if the file is stable and ready to be transferred. Stability may be determined by simple measures such as checking file size. Ingest pulls files, as they are available. If a data set consists of more than one file, ingest determines if all files that compose the data set are available before asserting that the data set is complete and is ready for cataloging, and archiving, and check the integrity of the data set based on the agreement between NESDIS and the data the provider.

2. To complete the reception of a data set, the Archive, Access and Distribution System names the data set with a unique name. The system employs a convention that assures a unique name for the data set. The original data set name is preserved in the inventory catalog. The renamed dataset is then forwarded to the Archive system for archival in the near-line tape robotic library.

3. Ingest obtains descriptive data from the data set, as provided by the data provider, to create a catalog entry for the data set. Ingest inserts the entry into the inventory catalog. Descriptive data provides search criteria that can be used during order placement.

4. In the secure area, ingest determines if the file meets the criteria for a subscription order. If it does, the order is placed with the distribution sub-system.

5. If applicable to that data provider, the system notifies the provider via email that the Archive, Access and Distribution System has the data set and verifies that the files of the data set were successfully ingested.

2. Order placement

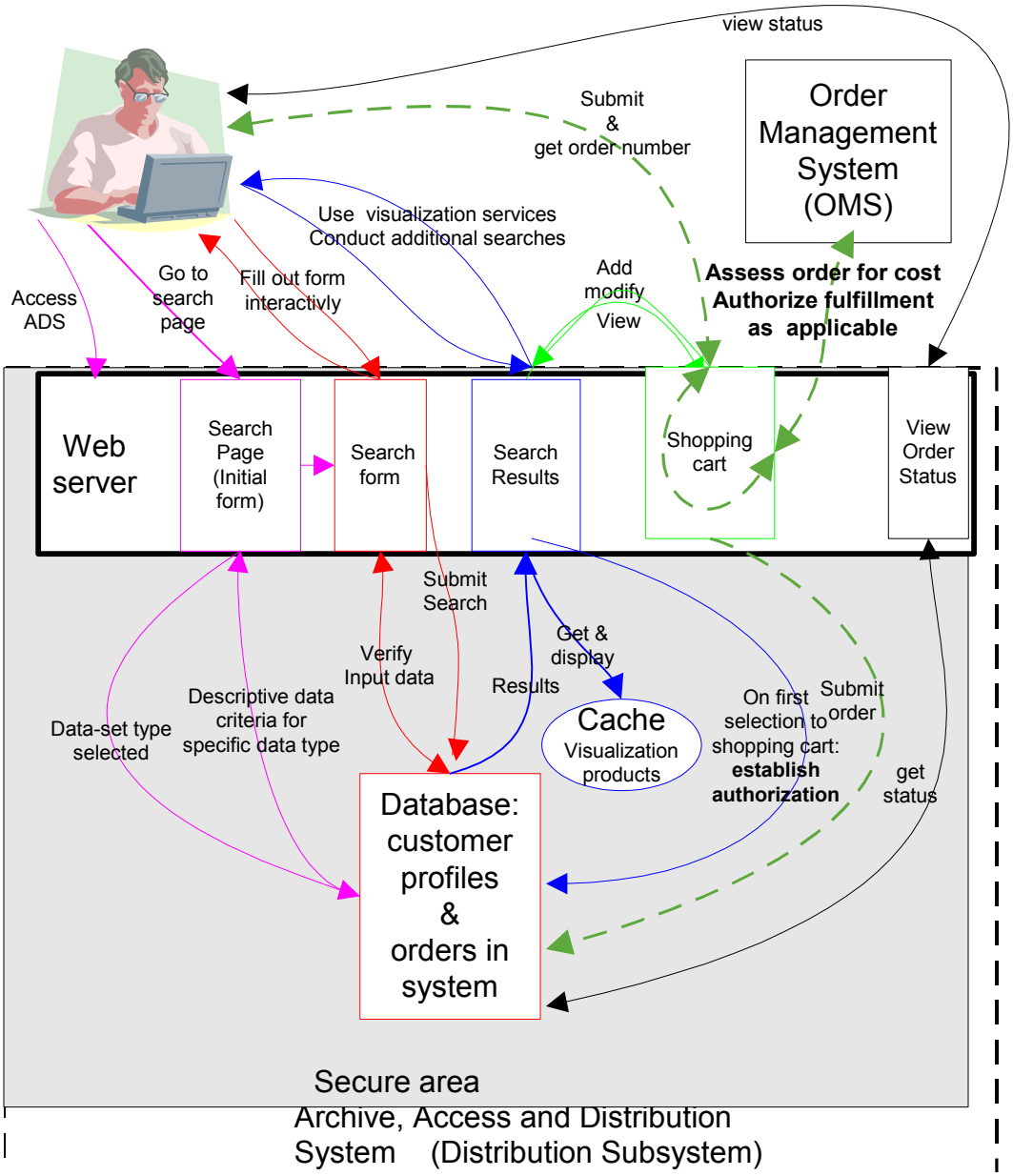

Figure 6 - Order Placement - Scenario 2

1. A customer, using a browser, accesses the Archive, Access and Distribution System via the Internet. The customer proceeds to the search page and starts filling out the search form for a new search. The customer may select a previous search form and modify the form. The form requires that a type of data be selected first. Then the form is configured to display options for descriptive data that is unique to that data type. Descriptive data includes geo-spatial, temporal information, and the origin of the data.

2. The customer completes the search form and submits the search.

3. For some data sets, visualization products are available to aid in the selection of products, as defined by agreement between NESDIS and the data provider.

4. Upon the first selection to add a file to the shopping cart, the system requests a password for authorization to create an order. The customer completes authorization protocols. Now, one or more data sets can be added to the order, and the customer can select delivery options. At this point, the customer may conduct additional searches for different types of data and add more datasets to the shopping cart or may submit the order. On the submittal of the order, if any of the data products of the order require customer charges to continue processing, the Order Management System's protocol to authorize the order is carried out. On completion, the Archive, Access and Distribution System accepts and submits the order to the delivery sub-system.

5. The customer is shown the number of the submitted order and, when the order processing begins, receives a confirmation email. The customer may view the status of the order at any time.

3. Delivery

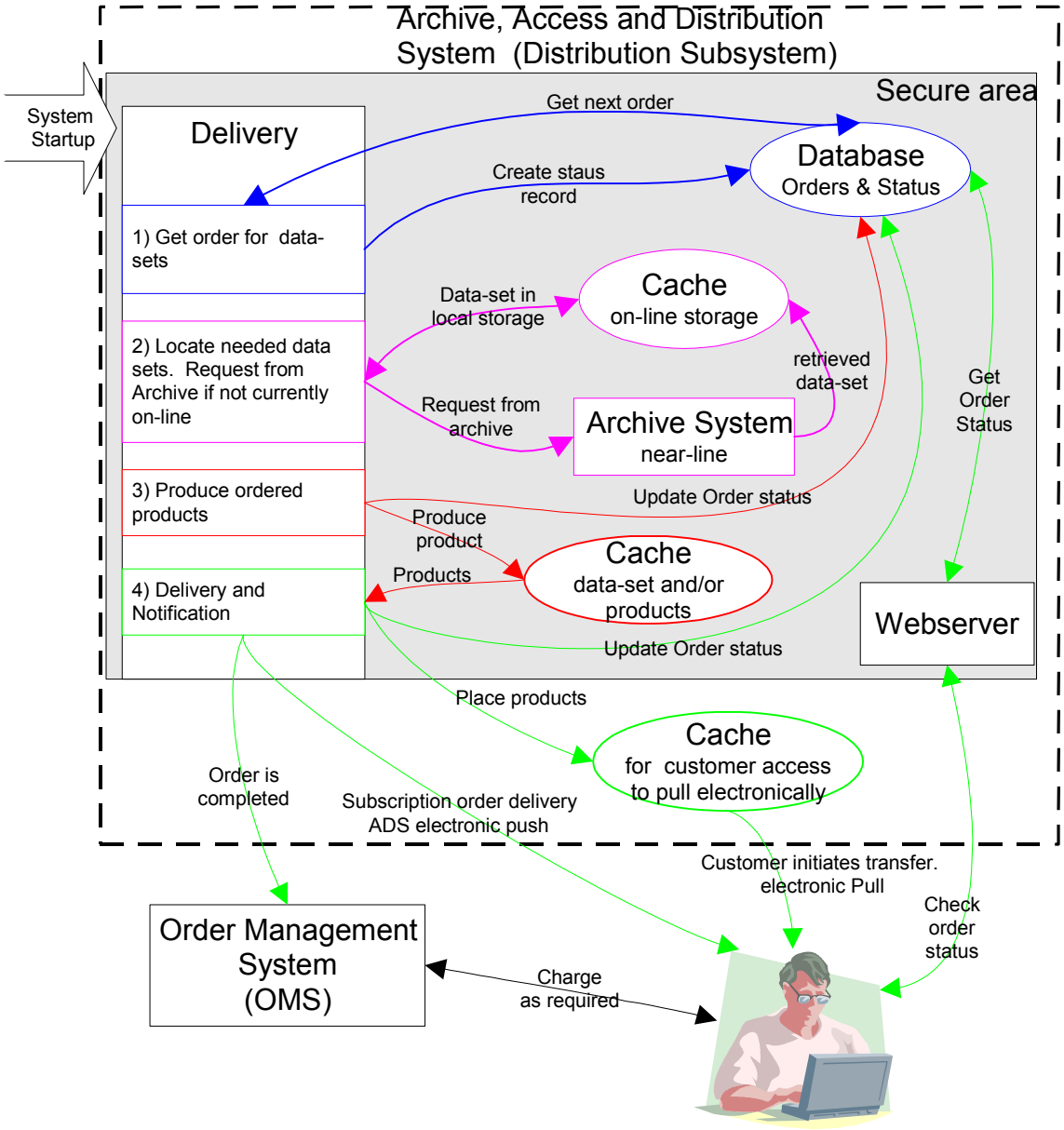

Figure 7 - Delivery - Scenario 3

1. The delivery sub-system selects the next order by priority.

2. First, the delivery sub-system determines if the data set resides in cache. The data set may still be in local disk cache if the data set is recent. Otherwise, the delivery sub-system requests that the archiving system retrieve the data from near-line storage. When retrieval from near-line storage is completed, the delivery sub-system is notified that the data set has been recalled and that processing is ready to continue.

3. If necessary, the delivery sub-system processes the data set to fulfill the user's request. Such additional processing might be required to sub-set the data set, change the format, compress or encrypt the data, etc.

4. The system completes delivery to the customer based on the customer's delivery options. A product for delivery is placed in an accessible location for the customer to obtain it by electronic transfer. A product ordered through a subscription may be pushed to the customer by electronic transfer. If attempts to push the data set fail, the subscriber is notified that the data set is available to be pulled. Other notifications to the customer are provided as options.

4. Customer support

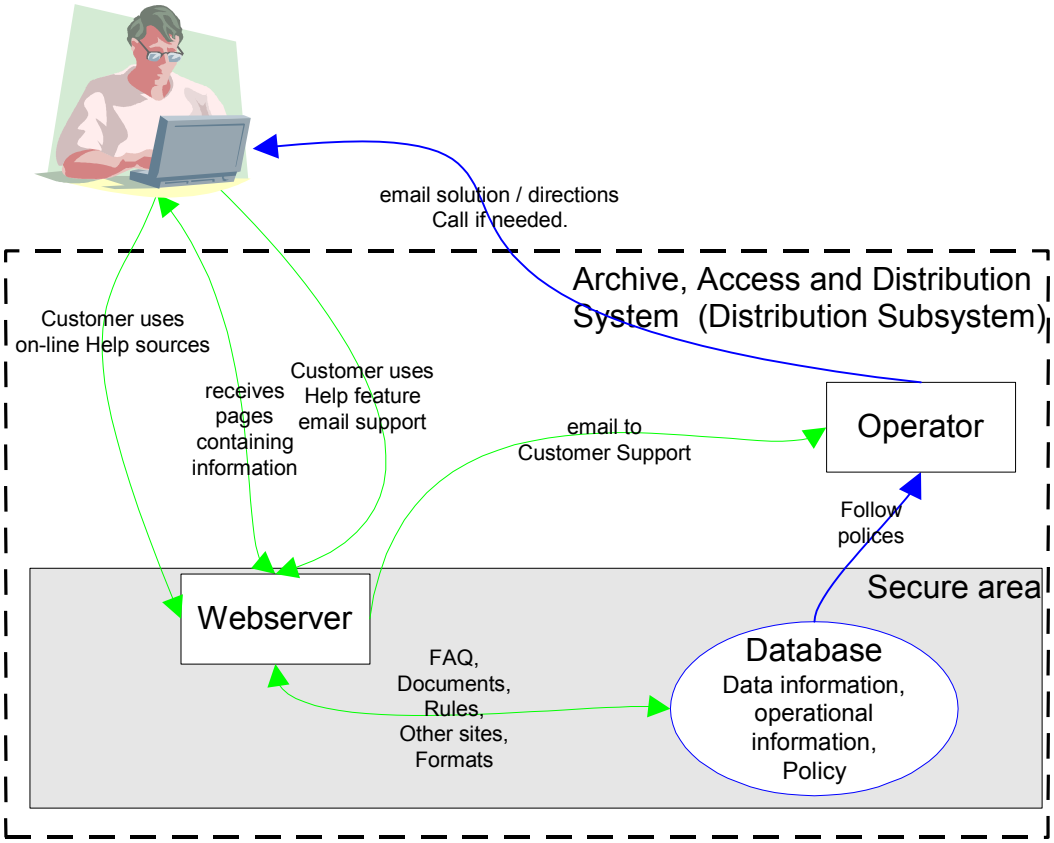

Figure 8 - Customer support - Scenario 4

1. A customer uses the on-line help sources to obtain information or sends an email to the Archive, Access and Distribution customer support.

2. An operator reads the help desk email inquiries and responds. The operator can satisfy the customer request by following the procedures set by the CPMT. Some of the activities that operators may perform in response to user email are: reporting problems with the user interface, providing documentation, investigating an order, and resetting a password.

6.5 Personnel Requirements

Operators need to have basic knowledge of Unix commands, system support, and moderate Structured Query Language (SQL) skills. Many status indicators and control parameters are contained in a database. Therefore, SQL skills are required for monitoring and controlling the system.

Customers need to be able to use a browser.

List of Acronyms

CIO	Chief Information Officer
CLASS	Comprehensive Large Array-data Stewardship System
COG	CLASS Oversight Group
CPMT	CLASS Project Management Team
DOD	Department Of Defense
DMSP	Defense Meteorological Satellite Program
EOS/MODIS	Earth Observing System/Modern Operations Distributed Infrared Spectrometer
GOES	Geostationary Operational Environmental Satellites
IJPS	Initial Joint Polar-orbiting Operational Satellite System
IPD	Information Processing Division
METOP	European Meteorological Operational Satellite
NASA	National Aeronautics and Space Agency
NCDC	National Climatic Data Center
NESDIS	National Environmental Satellite, Data, and Information Service
NEXRAD	NEXt generation weather RADAR
NGDC	National Geophysical Data Center
NNDC	NOAA National Data Centers
NOAA	National Oceanic and Atmospheric Administration
NPOES	National Polar-orbiting Operational Environmental Satellite
NPOESS	National Polar-orbiting Operational Environmental Satellite System
NPP	NPOESS Preparatory Program (NPP)
NVDS	National Virtual Data System
OMS	Order Management System
OSDPD	Office of Satellite Data Processing and Distribution
POES	Polar-orbiting Operational Environmental Satellites
RDBMS	Relational Database Management System
SAA	Satellite Active Archive
SQL	Structured Query Language